

Martin Hilpert

From *hand-carved* to *computer-based*: Noun-participle compounding and the upward strengthening hypothesis

Abstract: This paper addresses the recent history of noun-participle compounding in English. This word formation process is illustrated by forms such as *hand-carved* or *computer-based*. Data from the COHA shows that over the last two-hundred years, such forms have undergone a substantial increase in type and token frequency. These quantitative changes motivate an exploration of the qualitative changes that have accompanied them. A diachronic analysis of the noun and participle types that are recruited into noun-participle compounds reveals that the word formation process has changed substantially with regard to its component parts. While the observable changes in noun-participle compounding match several defining criteria of grammaticalization, it is argued here that these developments are more usefully seen as a process of constructional change. To distinguish between those two, this paper develops an idea that is called the upward strengthening hypothesis. According to this hypothesis, grammaticalization happens when the activation of a node in a constructional network strengthens not only that node itself, but also a node that is situated at a higher, more abstract level of that network.

Keywords: compounding, noun-participle compounding, grammaticalization, constructional change, Construction Grammar, corpus, COHA, upward strengthening hypothesis

DOI 10.1515/cog-2014-0001

Received April 15, 2014; revised September 1, 2014; accepted September 19, 2014.

1 Introduction

This paper addresses the recent history of noun-participle compounding in English. Forms such as *hand-carved* or *computer-based* are compounds that

Martin Hilpert: Institut de langue et littérature anglaises, University of Neuchâtel.
E-mail: martin.hilpert@unine.ch

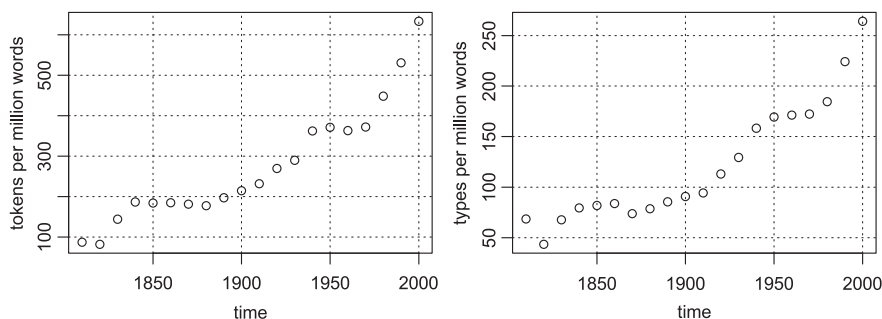


Fig. 1: Increases in normalized token and type frequencies in COHA

combine a noun and a past participle. This compounding strategy is recognized in overview works on English word formation and compounding (Fabb 2001: 68, Plag 2003: 153, Bauer 2006: 490, Bauer et al. 2012: 470). Major descriptive grammars of English (Biber et al. 1999: 534, Huddleston and Pullum 2002: 1659, Quirk et al. 1985: 1577) present it as a highly productive word formation process that appears to be related to the passive voice: A phrase such as *a government-funded project* can be paraphrased with a passive sentence such as *The project has been funded by the government*. The starting point of this paper is the observation that noun-participle compounding has recently undergone a substantial increase in token and type frequency. Data from the COHA (Davies 2010) shows that noun-participle compounds are much more common in modern texts and that they also occur in a greater variety of types. Figure 1 visualizes those increases.

Given these increases, it appears that something has been happening with noun-participle compounding, and it is the aim of this paper to explore what that is. Three leading questions will inform the discussion. The first question is how noun-participle compounding has changed with regard to its component parts. Word formation processes typically exhibit a number of constraints that restrict the formation of new types. Increases in type frequency can reflect a loosening of those constraints, so that speakers produce new coinages with lexical material that was previously unsuitable for the respective word formation process. To investigate whether this is the case in the present example, data from the COHA will be examined with regard to shifts in the lexical material that is recruited into noun-participle compounds. As will be discussed in more detail below, the data indeed reveal substantial changes, but these, in turn, do not reflect a systematic semantic broadening of the construction.

The second question is how noun-participle compounding relates to the passive. It has been mentioned above that the two constructions are often presented

as being mutually associated. Huddleston and Pullum note that forms such as *drug-related*, *home-made*, *safety-tested*, and *taxpayer-funded* “generally correspond to syntactic passives with a PP: *related to drugs*, *made at home*, *tested for safety*, *funded by taxpayers*, etc.” (2002: 1659). This list of examples already illustrates that instances of noun-participle compounding do not always correspond to a canonical long passive with a *by*-phrase. It is thus not always the agent of an action that is verbalized in the noun of a noun-participle compound, other semantic roles are possible. The research question that results for this study is how the relation between the two grammatical patterns can be best described. Lieber (1983) proposes that both the passive and noun-participle compounding are subject to restrictions that result from the argument structure of the past participle, which forms part of both the passive and noun-participle compounds. This proposal will be taken up here, but it will be re-conceptualized in terms of Construction Grammar (Goldberg 1995, 2006), specifically through the notions of inheritance and subpart links. For the purpose of the analysis, and following the work of Booij (2010) on construction morphology, this paper views both the passive and noun-participle compounding as constructions, that is, as form-meaning pairings that form part of speakers’ knowledge of language. Both constructions involve the past participle as a part of their morphosyntactic structure. Constructions that show an overlap in their structural components are said to be connected through a subpart link (Goldberg 1995: 78). The term inheritance (Kay and Fillmore 1999: 7) signifies that both constructions exhibit behavioral traits that they ‘inherit’ from the past participle, which is roughly comparable to living beings inheriting genetically transmitted traits from their ancestors. What will be explored in this paper is whether the observed changes in noun-participle compounds are paralleled by changes in the passive, and thus due to common inheritance, or whether these changes are in fact independent developments. Also this question is addressed through a corpus analysis of collocating lexical material. The main result of that analysis is that noun-participle compounding has been developing along a trajectory that is independent of the passive.

The first two questions lead up to a third one that probes the theoretical implications of the empirical results. What is at issue is whether the observed changes can be seen as an instance of grammaticalization. The developments that noun-participle compounding underwent match many of the criteria that are commonly associated with grammaticalization, understood here as “*the change whereby lexical terms and constructions come in certain linguistic contexts to serve grammatical functions, and, once grammaticalized, continue to develop new grammatical functions*” (Hopper and Traugott 2003: 18). As was shown in Figure 1 above, noun-participle compounding increased dramatically in frequency. An increase in token frequency strengthens the cognitive entrenchment

of a construction (Bybee 2006); greater type frequency relates to the phenomenon of host-class expansion, which Himmelmann (2004) views as a sign of grammaticalization; and the greater range in noun and participle types might be an expression of the construction widening its meaning spectrum through semantic bleaching (Sweetser 1988). Given that fairly diverse diachronic processes are routinely subsumed under the heading of grammaticalization (cf. the scope of grammaticalization phenomena in part IV of Narrog and Heine 2011), it is not immediately apparent why the development of noun-participle compounding should be excluded out of hand. This paper will however present an argument for an alternative view, drawing on concepts from Construction Grammar, specifically the ideas of constructionalization (Traugott and Trousdale 2013) and constructional change (Hilpert 2013). The core of the argument involves the cognitive process of schematization (Tomasello 2003: 122), i.e. the formation of abstract schemas on the basis of more specific tokens. What will be proposed here is that grammaticalization and constructional change are both processes that change a network of constructional knowledge, but that they do so in different ways: Whereas in grammaticalization, the experience of a linguistic unit leads to the progressive entrenchment of a more schematic construction, situated at a higher level in the constructional network, constructional change can manifest itself in the strengthening of several more specific sub-schemas, at lower levels of the constructional network. This proposal will be called the upward strengthening hypothesis. It will be argued that the developing network of noun-participle compounds, despite increases in types and tokens, does not show upward strengthening, and hence does not instantiate grammaticalization.

The remainder of this paper is structured as follows. Section 2 briefly surveys what has been said about noun-participle compounding, and it reviews Lieber's (1983) account of how that construction relates to the passive. It is explained how that relation can be modeled in Construction Grammar. The section further lays out two alternative hypotheses that follow from that account and that guide the subsequent empirical analysis. Section 3 discusses what data has been retrieved from the COHA for the present study. Following that, the empirical results will be presented and contextualized with the hypotheses from section 2. Section 4 develops the theoretical contribution of this paper, namely the question how the difference between grammaticalization and constructional change can be operationalized in terms of measurable developments in constructional networks, notably the entrenchment of an overarching constructional schema vs. the entrenchment of multiple subschemas. Section 5 concludes the paper with a few pointers for future research.

2 Noun-participle compounding and the passive

2.1 Introducing noun-participle compounds

Noun-participle compounds such as *market-oriented* exhibit distributional traits of adjectives. They can be used attributively (*the market-oriented approach*) and predicatively (*the approach is market-oriented*), in some instances they are gradable (*Current approaches are more market-oriented*) and adverbally modifiable (*the approach is strongly market-oriented*). It is also possible to coordinate a noun-participle compound with an adjective (*efficient and market-oriented approaches*). The adjectival qualities of noun-participle compounds find a natural explanation in the ambicategorical status of participles, which are known to combine verbal and adjectival features (see for instance Huddleston and Pullum 2002: 78). Participles such as *prepared* function as verbs in the perfect (*I have prepared a little snack*) or the passive (*Snacks will be prepared*); they function as adjectives in attributive constructions (*Pour the mixture into the prepared bowls*) or contexts such as the comparative correlative (*The more prepared you are, the better*). Since noun-participle compounds can be seen as right-headed compounds, it is to be expected that its instantiations inherit some of the distributional characteristics of the participial head.

Plag (2003: 153) points out that the adjectival qualities of noun-participle compounds may either result from the inclusion of the participle as a deverbal adjective, or from the affixation of the suffix *-ed* to a phrasal unit or a compound. As examples (1a) and (1b) illustrate, these two strategies result in different internal bracketing structures. In (1a), the suffix forms a constituent with only the verb, in (1b) the noun-noun compound *glass bottom* forms a constituent that is suffixed with *-ed*.

- (1) a. [market – [orient – ed]]
 b. [[glass bottom] – ed]

Examples that conform to the bracketing seen in (1b) usually carry applicative meaning, in this case ‘being equipped with a glass bottom’. Many common coinages of this type include body part nouns, as in *barrel-chested*, *wire-haired*, *eagle-eyed*, or *hawk-nosed*. Both bracketing types will be included in the present analysis, but the theoretical argument chiefly addresses the first one, which is overwhelmingly more frequent in both type and token frequency.

A further distinction between different types of noun-participle compounds is pointed out by Bauer et al. (2012: 470), who differentiate between argumental

compounds and non-argumental compounds. In an argumental noun-participle compound, the noun represents a core argument of the participial verb form. For instance, the form *doctor-recommended* is an argumental compound because the noun *doctor* would be understood as the agent who *recommended* a certain treatment. Core arguments can also be realized as prepositional objects. For example, the verb *relate* occurs in noun-participle compounds such as *drug-related*. In such coinages, *relate* instantiates the meaning of ‘have some connection with, have a relation to’ (OED sense 6a). Verbal uses of this sense project an argument in the form of a prepositional phrase with *to*, as in *One violent crime in five is related to drug use*. A corresponding sentence without that argument would be judged as incomplete (**One violent crime in five is related*). Importantly, not all instances of noun-participle compounding involve arguments. An example of a non-argumental coinage would be *home-cooked*. This form corresponds to a sentence such as *The food was cooked at home*, which readily allows for the omission of the prepositional phrase (*The food was cooked*). In non-argumental examples of noun-participle compounding, the noun may express a location (*Texas-born*), a cause (*honey-flavored*), a purpose (*safety-tested*), or an instrument (*pencil-written*), amongst other possibilities.

While this structural and semantic variety suggests that noun-participle compounding is a relatively unconstrained word formation process, Bauer et al. (2012: 470) observe that it is restricted in one fundamental way: The noun in a noun-participle compound cannot represent the direct object of a transitive verb. Hence, potential forms such as **lunch-eaten*, **door-locked*, or **book-read* are ruled out. In the following, this regularity will be referred to as the no-object constraint. A few apparent counterexamples to the no-object constraint are shown in (2). For each of these, the noun could be seen as the direct object of the verb that appears in the participial form.

- (2) beer-drunk, corn-fed, throat-cut, wing-broken, calcium-added

The coinage *beer-drunk* is perhaps easiest to explain away, since here the form *drunk* is clearly not the participle of *drink*, but instead a derived adjective that means ‘intoxicated’. For *corn-fed*, *throat-cut*, and *wing-broken* the case could be made that simple transitive sentences are not the only possible correspondences. Animals can be *fed with corn* or *cut through the throat*; *break* is an ergative verb which also yields *The wing broke* as a grammatical sentence. More problematic is the case of *calcium-added*, since the verb *add* is very much restricted to transitive argument structure. It is easily possible to find further formations with *added* on the internet (*iron-added*, *vitamin-added*, etc.), so the anomaly of this case is to be taken seriously. A possible explanation for it relates to the phrase

no sugar added, which is a very common product description that even occurs in attributive contexts (*no-sugar-added apple sauce*). The form *calcium-added* may have come about through analogy with such usage patterns, but this is a speculative suggestion that will not be further discussed here. For now, it appears that the no-object constraint is remarkably robust and thus in need of an explanation. The next section will discuss one such explanation, which at the same time illuminates the connection between noun-participle compounding and the passive.

2.2 Towards an explanation of the no-object constraint

Lieber (1983) investigates why potential compounds such as **during-happy*, **carefully-considering*, or **strawberry-picked* are unacceptable and uninterpretable to speakers of English. This is not a trivial problem since compounding as such is a highly productive and relatively unconstrained way of generating new English words. Lieber's suggestion involves a syntactic concept that is not only at work in compounds, but in fact in many other grammatical and lexical structures. What she proposes is that compounding is constrained by the respective argument structures of the component elements. Argument structure, often also called valency, is most prototypically seen as a characteristic of verbs (Hilpert 2014: 26). A verb such as *give* projects three arguments, namely someone doing the giving, something that is given, and someone who receives the given object. The verb *yawn* only projects a single argument, namely the agent of the yawning. Like verbs, also other parts of speech have argument structure. For example, a preposition such as *in* projects an argument, namely the landmark location in which some other thing is located. By contrast, there are many lexical elements that do not project any arguments, notably nouns such as *sky* or *truck* and adjectives such as *wild* or *poor*. Lieber's proposal correctly predicts that compounding with nouns and adjectives, which are free of argument structure constraints, is rampantly productive. It also correctly predicts that the systematic gaps that can be observed in the inventory of English compounds should involve verbs and prepositions, that is, elements that project arguments. Table 1 shows an overview of primary compounding types that is offered by Plag (2003: 143). As is apparent, all gaps and problematic types involve either a verb, a preposition, or both. It could be added that the verb-verb compound *stir-fry* forms part of a very small set (including *freeze-dry* and *drop-kick*, but few additional coinages), so that yet another question mark might have been in order.

Given this initial support for the idea that argument structure restricts the productivity of compounding, how does Lieber envision the constraints on

Table 1: Inventory of primary compound types (Plag 2003: 143)

	Noun	Verb	Adjective	Preposition
Noun	film society	brainwash	knee-deep	–
Verb	pickpocket	stir-fry	–	breakdown (?)
Adjective	greenhouse	blackmail	light-green	–
Preposition	afterbirth	downgrade (?)	inbuilt (?)	into (?)

compounds with verbs and prepositions? Lieber (1983: 259) states that there is an asymmetry between the left and the right element of a compound. If the left element projects an argument, that argument must be the right element of the compound. This is the case in *pickpocket* and *afterbirth* from the table above, but already the compound verb *stir-fry* goes against that generalization, as *fry* is not an argument of *stir*. Generally, compounds that are predicted to be impossible according to this account are forms that begin with a verb or preposition and that have a right element that does not correspond to an argument of the left element. Some such forms are shown below (from Lieber 1983: 261–262).

(3) *seem-dog, *hit-happy, *among-legal, *during-warm

If the right-hand element of a compound projects an argument, the consequences are different. The right-hand element acts as the head of a compound and hence determines its part-of-speech category and its argument structure. If the right element projects arguments, those arguments may thus be found outside of the compound itself. Lieber (1983: 258) offers the example of *handweave*, in which *weave* projects an object that is found in a separate noun phrase, outside the compound itself, in a sentence such as *They were handweaving cotton rugs*. The only constraint on such compounds is that the left-hand element must be semantically relatable to the right element in some way. This allows *handweave*, but bars forms such as **blue-shave*.

Importantly, Lieber’s proposal extends to compounds in which the right-hand elements are deverbal, as is the case in noun-participle compounds such as *hand-picked*. Two points are of particular relevance. First, it follows from what has been said above that the overall noun-participle compound has the participle as its head and thus projects the same arguments as that participle. Second, Lieber argues that past participles such as *picked* or *given* have argument structures that deviate systematically from the verb stems *pick* and *give*. Her elaboration of this point is worth quoting in full:

Essentially, the argument structure of the passive participle is viewed here as an operation on the argument structure of the verb stem to which the participle is related; this operation removes the original external argument (most often the Agent) and makes the internal argument the new external argument. This captures the standard observations about the relationship between active and passive sentences, namely, that the object of the active becomes the subject of the passive and that the subject of the active is no longer an obligatory argument of the passive. (Lieber 1983: 274)

The quote evokes theoretical concepts of transformational grammar which need not concern us here. According to the explanation, the no-object constraint of noun-participle compounding (**lunch-eaten*) and the promotion to subject that happens in the passive are actually two expressions of the same underlying phenomenon. Lieber suggests that a transitive verb such as *pick* has an argument structure that includes an internal argument, a patient, and an external argument, an agent, yielding sentences such as *The boy picked strawberries*. In the past participle *picked*, the argument structure is different. The original external argument (*the boy*) is not part of the argument structure, which merely specifies a ‘new’ external argument (*strawberries*). The two argument structures are shown in (4).

- (4) a. *pick* (internal: PATIENT, external: AGENT)
 b. *picked* (external: PATIENT)

The consequence of this is that constructions that involve a past participle (*picked*) have to realize the patient argument (*strawberries*) outside of the syntactic constituent that is headed by the participle. In the passive, that means outside the verb phrase, leaving only the subject noun phrase as a possible site; in noun-participle compounding, that means outside the compound itself, either in an attributive or in a predicative construction. These structures are illustrated in (5).

- (5) a. The strawberries [were picked]_{VP}.
 b. I like [hand-picked]_{ADJ} strawberries.
 c. The strawberries were fresh and [hand-picked]_{ADJ}.

Lieber’s account is thus able to explain why direct objects of transitive verbs are the only structures that are reliably barred from noun-participle compounds, while other structures, as long as they have some semantic relation to the participle, can be flexibly integrated. The account furthermore clarifies the relation between noun-participle compounding and the passive. Both include the past participle, which can be thought of as an argument structure construction that ‘expels’ the object of a transitive verb.

In what follows, that analysis will be adopted in spirit, but slightly changed with regard to its theoretical notation and its reliance on transformative operations. In order to describe relations between different morpho-syntactic constructions, Lieber uses the term feature percolation (1983: 252). Amongst other things, this term describes the phenomenon that a noun-participle compound shares its behavioral traits with the participle that is the compound head. The features of the participle are thought to ‘percolate upwards’ to the syntactic node that dominates the entire compound. Feature percolation has a direct counterpart in the theoretical machinery of Construction Grammar, where the term ‘inheritance’ evokes a slightly different metaphor to express the same phenomenon. The term signifies a relation between two constructions, such that certain aspects of form and/or meaning are shared between them. Kay and Fillmore (1999: 7) offer the example of the transitive verb phrase construction, which inherits a formal aspect from the more abstract head-complement construction, namely the relative order of head (verb) and complement (object). This type of inheritance is called an instance link (Hilpert 2014: 60), since the verb phrase construction instantiates the head-complement construction. Transferred to the case of noun-participle compounding, we can say that the past participle construction is instantiated in noun-participle compounding. The same inheritance link connects the past participle with the passive. In each passive, we find an instance of the past participle. Instantiation links of this kind are essentially a notational variant of what Lieber calls feature percolation. Constructions are however not only related through instantiation links. A type of link that is important to the present analysis concerns the relation between noun-participle compounding and the passive. These two do not instantiate one another, rather, they share a subpart of their respective structures. In Construction Grammar, that kind of relation is captured in a so-called subpart link (Hilpert 2014: 62). Other pairs of constructions that are related in this way are the transitive and the ditransitive construction (*John wrote a letter* / *John wrote Mary a letter*), or the passives with *be* and *get* (*John will be fired* / *John will get fired*). What a subpart link implies is the assumption that in speakers’ knowledge of language there is a connection between those constructions.

To summarize the main point of this section, the no-object constraint in noun-participle compounding can be explained, pace Lieber (1983), but couched in a constructional framework, through the inheritance of features from the past participle, most notably its argument structure. That argument structure requires a realization of the patient argument outside of the compound itself. This characteristic is shared with the passive, where the same argument structure constraint leaves the patient in the position of the subject, outside of the verb phrase.

The next section explores what all of this implies for the diachronic analysis of noun-participle compounding.

2.3 Two hypotheses

The fact that both constructions inherit aspects of their argument structure from the past participle allows the formulation of two conflicting hypotheses that can be empirically tested. According to the first of these, speakers treat noun-participle compounding and the passive as ‘the same’ with regard to their combinatorics. The two constructions share a fundamental constraint, and this could lead speakers to the assumption that in fact all of their constraints should match. Treating the two constructions in this way would come with the benefit of storage economy. Speakers would not have to memorize two different behavioral profiles, but could simply extrapolate. If they have observed one construction, they know how the other one works. Alternatively, it could be that speakers treat the two as separate constructions. Both constructions inherit characteristics from the participle, but beyond that, each construction has its own behavioral profile. This would be less economical, but perhaps motivated in terms of the different communicative functions that the two constructions serve.¹

With regard to the empirical analyses in the next section, the first hypothesis predicts that any collocational shifts that can be observed with noun-participle compounding are paralleled by changes in the passive. We can allow for a certain amount of variability that might reflect free variation or sampling error in the corpus data, but there should not be substantial, measurably significant differences. Conversely, such differences are predicted by the alternative hypothesis. The observation that noun-participle compounding exhibits a unique set of collocational preferences would support the idea that speakers mentally represent this word formation process as a construction, a linguistic unit with characteristics that are not reducible to more general linguistic patterns.

¹ An anonymous reviewer asks why the passive is the only construction that is singled out for a comparison with noun-participle compounding. It is a valid point that noun-participle compounding can be thought to entertain inheritance relations with other constructions such as adverbial-participle compounds (*oddly-shaped*) or pseudo-participle compounds (*high-walled*). Yet, the specific claim to be investigated here is whether the passive and noun-participle compounding need separate mental representations or not. Such a comparison can be carried out without reference to other constructions.

3 Data, methodology, and results

3.1 Data retrieval

The data for this study is taken from the Corpus of Historical American English (COHA, Davies 2010), which is a diachronic corpus that represents American English from the 19th to the 21st century with about 400 million words of running text. The corpus thus allows comparisons between 20 decades of language use. The amounts of text and the relative proportions of different genres are not exactly matched, so that later decades include more data. Especially the first few decades are relatively underrepresented. The procedure for data retrieval was the following. The COHA interface was queried for forms that contained a hyphen, were tagged as adjectives, and that ended in a letter qualifying as the last letter of a past participle, chiefly *d* (*turned, cooked*), *t* (*built, cut*), and *n* (*broken, written*), but also *e* (*gone, made*), *g* (*hung, stung*), *k* (*struck, sunk*), and *m* (*swum*). These procedures ensured high recall, but incurred low precision. Forms such as *able-bodied*, *across-the-board*, or *ad-libbed* had to be manually excluded. A reasonable question to ask is whether an exclusive reliance on hyphenated forms is justifiable, since many English compounds are juxtapositions (*film society*) or univerbations (*greenhouse*). One might expect that particularly frequent combinations show a tendency towards univerbated orthography and that less frequent combinations vary between juxtaposition and hyphenation. Table 2 shows four noun-participle compounds and their COHA frequencies in the three orthographic variants. The table reveals that the hyphenated forms are by far the most frequent option for these four coinages, but that juxtaposed and univerbated forms in these cases account for more than 10% of the data, which is too large a margin of error.

To close in on that margin (which is even larger for other coinages, see below), the 1000 most frequently attested hyphenated types were used to search for their juxtaposed and univerbated counterparts. The results were added to the existing database. The largest additions in terms of frequency were idiomatic univerbated forms such as *moonlit*, *heartfelt*, or *bloodshot*. While this procedure does not re-

Table 2: Orthographic variants of noun-participle compounds

	Hyphenation	Juxtaposition	Univerbation
State-owned	283	12	11
Hand-painted	225	13	26
Faith-based	100	9	2
Home-cooked	87	3	4

trieve additional types, it does assure that the token frequencies at the upper end of the spectrum are accurate. The final dataset includes 31,659 different types of noun-participle compounding. These types represent 3,200 different participles and 8,291 different nouns. Across the twenty decades from the 1810s to the 2000s, the token frequencies of those types add up to roughly 127,000 data points.

In order to compare the development of noun-participle compounding to that of the passive, a second dataset, representing the *be*-passive, had to be retrieved from the COHA. Here, an exhaustive retrieval could not be attempted for practical reasons, as the passive is realized in a number of different forms, some of which lack a form of *to be* (as in *Luggage left unattended will be destroyed*). As an approximation, the COHA was searched for past participles that were directly preceded by a form of *to be*. This disregards passives that are negated or that involve adverbial modification, which has to be kept in mind as a limitation. The search procedure retrieved 5,384 different participle types and close to 3 million tokens. Not all of the participles that are found with the passive are also found with noun-participle compounds, and vice versa. Between the two datasets, there is an overlap of 2,075 participle types. The full dataset, including descriptive statistics and R code for the analysis that is described below, is available from the author upon request.

3.2 Data analysis

It was shown in Figure 1 above that noun-participle compounding has undergone substantial increases in type and token frequencies over the past two centuries. Diachronic studies of word formation processes typically also consider the occurrences of hapax legomena, i.e. types that only occur once, and the ratio of those types to the grand total of all tokens, which is a common measure of morphological productivity (Baayen 2005: 244). In the database, the absolute number of hapaxes increases over time, but the ratio of hapaxes stays fairly constant around 30% (mean = 34.1%, SD = 7.8%). Since the sizes of the corpus periods vary, these figures should not be interpreted directly in terms of developing morphological productivity. What can however be said is that across all of the COHA, noun-participle compounding has been very productive, as would in fact be expected of a compounding word formation process.²

² An anonymous reviewer raises the question why the productivity of noun-participle compounding is not assessed in terms of measuring its success relative to other competing constructions. This could be done for instance by checking how noun-participle compounding contributes to overall vocabulary growth (Baayen 2005). This is a valuable suggestion that could unfortunately not be carried out due to limitations in the online interface of the COHA.

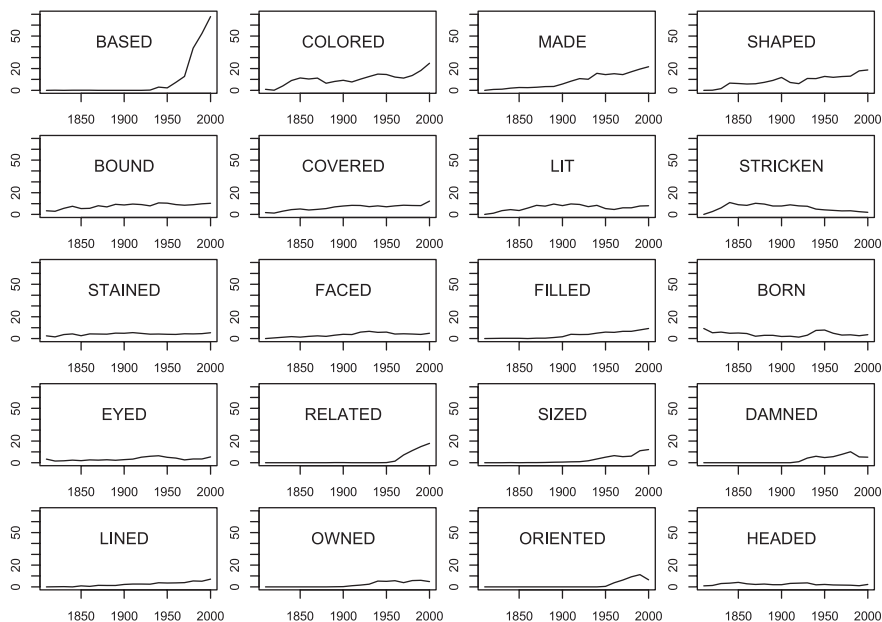


Fig. 2: Normalized token frequencies of the 20 most frequent participles in noun-participle compounds

3.2.1 Which nouns and participles carried the frequency increases?

A first question to explore on the basis of the dataset is what nouns and participles carried the frequency increases. Figure 2 shows how the 20 most frequent participles develop in terms of token frequency. The panels are ordered according to normalized token frequencies (tokens per million words), shown on the y-axis, drawn to the same scale. The x-axis shows the progression of time. What can be seen in the panels of the graph is that certain participles have undergone frequency increases that started already in the 19th century (*colored*, *made*, *shaped*, etc.), whereas others have to be seen as recent success stories, notably *based*, but also *related*, *sized*, and *oriented*. Amongst the participles shown, only *stricken* and *born* undergo decreases. The main finding to take away from this graph is that the participle *based* shows a unique development and accounts for a large portion of the recent frequency increases in noun-participle compounding.

It is instructive to compare the token frequency changes from Figure 2 to developments that concern type frequency, that is, family sizes of the participles. Figure 3 shows the 20 participles with the largest word families in the dataset. The y-axis shows types per million words. Many of the participles from Figure 2

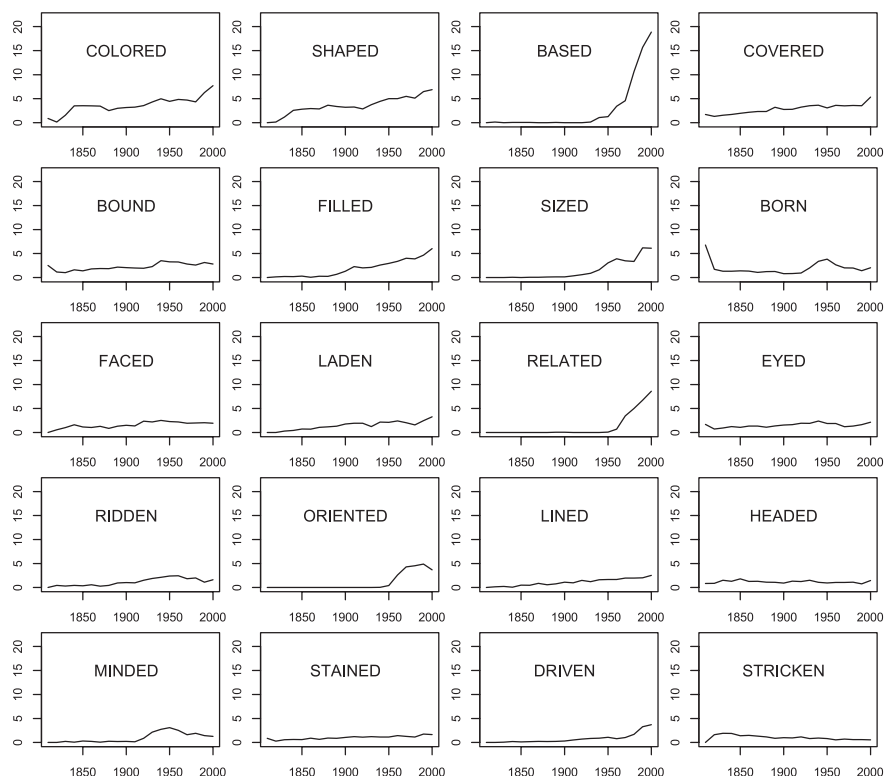


Fig. 3: Normalized type frequencies of the 20 largest participle families in noun-participle compounding

re-appear, but tellingly, *made*, *lit*, *damned*, and *owned* do not. These participles have relatively small families, often headed by a highly frequent, idiomatized element, such as *homemade*, *moonlit*, *goddamned*, or *state-owned*. Another observation that can be made on the basis of Figure 3 is that some participle families seem to represent linguistic fashions. The participle *minded*, in the lower left corner of the graph, serves as an example. Taylor (2012: 115–17) investigates compounds with that participle using the TIME corpus (Davies 2007), observing that many such forms show frequency developments with peaks in the mid-20th century. This observation is replicated here with data from COHA. Coinages with the participle *minded* are formed less often today than they used to be. Other participles in Figure 3 show similar developments. Coinages with *born* seem to have enjoyed a certain popularity around 1950; the recent success of *oriented* may already be on the decline.

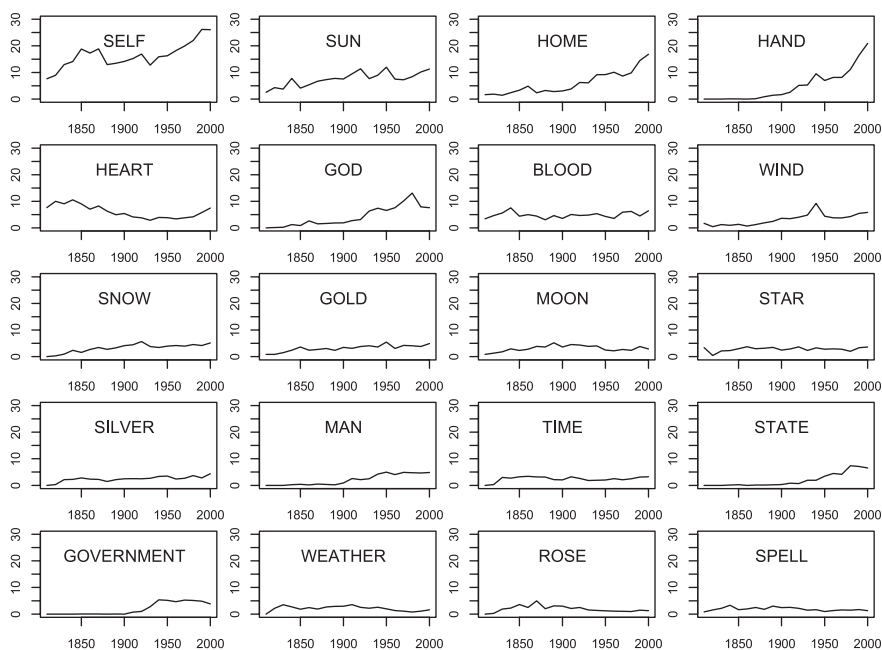


Fig. 4: Trends of the 20 most frequent nouns in noun-participle compounds

Figure 4 shows token frequency developments in the 20 most frequent noun types. Normalized token frequencies (tokens per million words) are seen on the y-axis, time is on the x-axis. Again, it is possible to distinguish long-term developers such as *self*, *sun*, and *home*, from more recent starters, which in this case are *man*, *state*, and *government*. However, in contrast to the participles and the case of *based*, there is no noun type that could be singled out as outdoing all others.

Coming back to the participle types, it is tempting to think that the overall frequency increase of noun-participle compounding reflects an ever-increasing pool of participles that are recruited into the construction and that thus reflect a process of semantic broadening. It will be argued in this paper that this idea is only partly true. Figures 5 and 6 track the developments of the 30 most frequent participle types in terms of normalized token frequency and compares those developments to the remaining 3,270 participle types. The grey-shaded areas on the upper side of the graphs represent the frequent participles; the large light grey area on the lower side of the graph represents the infrequent types. Figure 5 shows the development in absolute terms; Figure 6 represents the same development in terms of relative frequency.

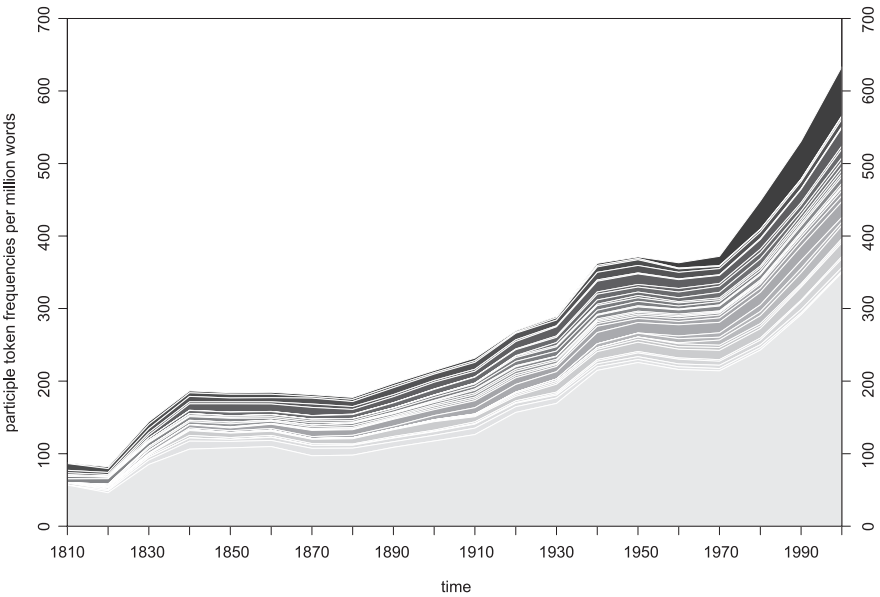


Fig. 5: Absolute frequency developments of 30 frequent particle types (top) and 3,270 infrequent particle types (bottom grey area)

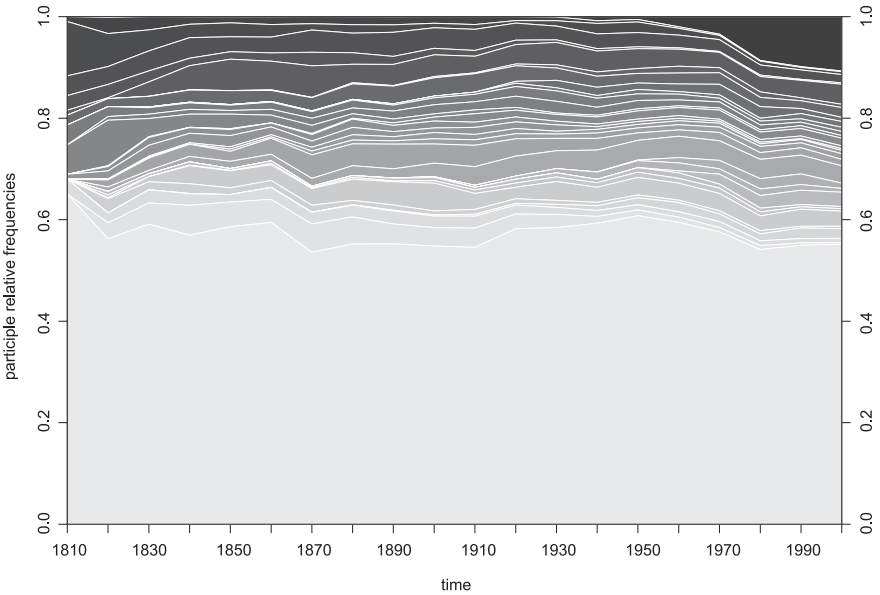


Fig. 6: Relative frequency developments of 30 frequent particle types (top) and 3,270 infrequent particle types (bottom grey area)

While it is not to be disputed that both frequent and infrequent participle types gain in absolute frequency, Figure 6 shows that in relative terms, the ‘upper crust’ of 30 highly frequent participle types has been consistently representing a share of about 40% of all usage events. What would be expected in a scenario of sustained semantic broadening is that low-frequency types account for a progressively larger ratio of tokens. This is not the case. The interpretation of Figures 5 and 6 that will be argued for here is that certain participle types, among them *based* (shown in black at the very top), *related*, and *colored*, establish themselves as highly productive mid-level schemas of the noun-participle compounding construction. Each of these participles accounts for some of the overall frequency increase of noun-participle compounding, but their respective successes do not necessarily trigger new coinages with other participles. In other words, the high type frequencies of NOUN-*based* and NOUN-*related* do not strengthen the mental representation of a more abstract noun-participle compounding construction that in turn would more easily license (unattested) forms such as *choir-sung* or *vinegar-cleaned*. This claim will be defended in more detail in section 4, where a number of testable predictions will be derived from it.

3.2.2 Developments in noun-participle compounding and in the passive

How do the developments in noun-participle compounding compare to those in the passive? For each of the 2,075 participle types that occur in both constructions, token frequency measurements from 20 successive decades are available. If the two constructions change in related ways, the respective trends in those measurements should correlate. If a verb increases in token frequency in the passive, it should also become more frequent in noun-participle compounding. Naturally, frequency decreases should also correlate. In order to test this idea, all paired vectors of passive frequency measurements and compounding frequency measurements were submitted to a correlation statistic (Pearson’s r). For a large majority of participles (79.6%), the results were non-significant. The mean of all returned coefficients, which lies at 0.16 (SD 0.31), is very small, the distribution around that mean is unimodal. This means that there is no positive evidence to support a putative relation between noun-participle compounding and the passive. However, can it be substantiated that there are robust differences between the two? In order to answer this question, a distinctive collexeme analysis (Gries and Stefanowitsch 2004) was carried out on the basis of the two datasets. A distinctive collexeme analysis compares how often a lexical item occurs across two alternative constructions. On the basis of the overall frequency of that element in the two constructions, the analysis determines with a Fisher exact test whether

Table 3: Participles that are strongly attracted to noun-participle compounding

Participle	Compounding		Passive	
	Observed	Expected	Observed	Expected
Bound	3,465	620.60	11,186	14,030.40
Based	5,046	607.17	9,288	13,726.83
Covered	2,917	472.13	8,229	10,673.87
Beaten	822	150.84	2,739	3,410.16
Related	1,433	169.82	2,576	3,839.18
Headed	966	145.88	2,478	3,298.12
Grown	933	139.36	2,357	3,150.64
Swept	793	133.26	2,353	3,012.74
Controlled	809	111.70	1,828	2,525.30
Owned	1,067	113.35	1,609	2,562.65
Faced	1,600	135.38	1,596	3,060.62
Lined	1,190	117.89	1,593	2,665.11
Possessed	690	94.25	1,535	2,130.75
Contained	671	85.69	1,352	1,937.31

the observed frequency is significantly higher or lower than the expected frequency. The analysis returns ranked lists of lexical items that are significantly attracted to either one of the two constructions. The null hypothesis that is refuted by significant results is that there is mere chance variation in how lexical items are distributed across the two constructions. A distinctive collexeme analysis of the COHA data establishes that the passive and noun-participle compounding are in fact substantially different in their collocational preferences. Tables 3 and 4 present some of the participles that show the most pronounced differences in their respective distributions.

In Table 3, several participle types from Figures 2 and 3 reappear, among them *bound*, *based*, *covered*, and *related*. All of these are frequently used in the passive as well, but while their relative frequency in noun-participle compounding is high, their relative frequency in the passive is much lower than expected. Table 4 reveals that many of the participles that are most distinctive for the passive are actually not used, or only very marginally so, in noun-participle compounds. Participles of common verbs such as *bring*, *do*, *find*, and *go* are not recruited into compound types. It would be an interesting question to pursue why speakers do not talk about *a weekend-done job* or *attic-found antiques*, which do not seem to be downright ungrammatical, but merely unusual. However, it will be left for future work to explore what constraints govern the coinage of new noun-participle compounds. The main point of this section has been to argue that

Table 4: Participles that are strongly attracted to the passive

Participle	Compounding		Passive	
	Observed	Expected	Observed	Expected
Given	335	1,661.96	38,900	37,573.04
Known	31	1,075.54	25,360	24,315.46
Used	11	1,472.70	34,756	33,294.30
Taken	6	1,810.04	42,725	40,920.96
Called	3	1,809.32	42,711	40,904.68
Considered	1	845.95	19,970	19,125.05
Expected	1	1,003.95	23,700	22,697.05
Gotten	1	758.82	17,913	17,155.18
Brought	0	1,056.31	24,937	23,880.69
Done	0	2,259.94	53,352	51,092.06
Found	0	2,140.74	50,538	48,397.26
Gone	0	1,357.78	32,054	30,696.22
Heard	0	771.82	18,221	17,449.18
Left	0	1,110.95	26,227	25,116.05

noun-participle compounding and the passive exhibit different behavioral profiles in terms of the participles that they host. The available evidence strongly supports the idea that speakers treat the two as separate constructions, despite the fact that they share a morphosyntactic component. With this observation in place, we can now move on to the third and final question of this paper, namely the question whether the changes that can be observed in noun-participle compounding constitute grammaticalization.

4 Grammaticalization or constructional change?

The starting point of this paper was the observation that noun-participle compounding has recently undergone substantial increases in type and token frequency. The empirical analyses in the preceding section established that the development of the construction has proceeded independently of developments in the passive, and that a sizable part of the recent frequency increases is represented by coinages with highly prolific participles, such as *based*, *related*, or *oriented*. All of this supports the idea that noun-participle compounding is currently undergoing a process of change. This section will discuss what type of change that process might be. Two contenders will be considered, namely grammaticalization on the one hand and constructional change on the other. The following paragraphs will present arguments for either position, but it will ultimately be

argued that the developments in noun-participle compounds instantiate constructional change.

4.1 Grammaticalization as a plausible contender

To start out, it is perhaps necessary to justify why the development of noun-participle compounds should even be considered as a case of grammaticalization. The phenomenon certainly does not instantiate the classic scenario of an independent lexical item developing into a morphologically bound, obligatory marker of a grammatical category (Lehmann 2002: 3–4 on Meillet). Word formation processes are rarely discussed as the product of grammaticalization (cf. the introductory discussion in Trips 2009: 1–5), and if they are, the discussion focuses on the initial emergence of the word formation process, not subsequent changes. Another weighty argument concerns the fact that throughout the time span that is considered in this study, noun-participle compounding has not undergone formal changes in terms of its morphosyntax or its phonological characteristics. It is maintained here that changes in frequency and combinatorics do show that change is going on, but the absence of formal change clearly detracts from the view that this change might be grammaticalization. That said, noun-participle compounding is a grammatical pattern, an argument-structure construction that shares structural traits with the passive, and so any developments that it undergoes will fall under broad definitions of the term grammaticalization. One such definition is offered by Wiemer and Bisang (2004: 4), who include “*all the processes involved in the diachronic change and the emergence of such [grammatical, MH] systems*”. Another recent definition of grammaticalization frames it as the change that gives rise to discursively secondary expressions (Boye and Harder 2012). The grammatical pattern that underlies concrete coinages such as *hand-carved* or *computer-based* is discursively secondary, as it cannot be itself the focus of an utterance. Hence, excluding the developments in noun-participle compounding from the domain of grammaticalization is not as straight-forward as one might imagine. While it is certainly possible to draw a line on the basis of some criterion, it should be remembered that different researchers will draw that line in different ways, and some not at all.

4.2 Grammatical constructionalization vs. constructional change

Research into grammaticalization has recently adopted a range of concepts from Construction Grammar (Noël 2007, Hilpert 2013, Traugott and Trousdale 2013,

amongst others). What is developed in these works is a re-conceptualization of language change in terms of changes that affect a structured network of constructions, which is meant to represent speakers' knowledge of language. Traugott and Trousdale (2013: 22) give special emphasis to the emergence of new nodes in such a network, for which they reserve the term *constructionalization*. In order for a linguistic unit to count as a new node, it has to have a new morphosyntactic structure and a new meaning. Constructionalization can happen instantaneously and create new words, in which case Traugott and Trousdale speak of *lexical constructionalization*, or it can happen gradually and create constructions that are procedural (i.e. discursively secondary in Boye and Harder's terms), which means that grammatical constructionalization takes place. The latter term comes close to being a synonym of grammaticalization, as understood against the theoretical background of Construction Grammar. Traugott and Trousdale (2013: 112) propose that grammatical constructionalization shows itself in three separate diagnostics. The first of these is an increase in productivity, usually accompanied by an increase in frequency of use. The second diagnostic is an increase in schematicity, which shows itself for instance in the relaxation of semantic constraints. Lastly, grammatical units that constructionalize tend to undergo a decrease in semantic compositionality, so that the meaning of the whole construction is no longer completely derivable from the meaning of its parts. Constructionalization, i.e. the emergence of a new node in the constructional network, is contrasted with constructional change, which Traugott and Trousdale (2013: 26) define as "*a change affecting one internal dimension of a construction*" that "*does not involve the creation of a new node*". The distinction between constructionalization and constructional change produces a clear verdict for the case of noun-participle compounding, which fails to show an increase in schematicity or a decrease in semantic compositionality. According to Traugott and Trousdale's criteria, the changes in noun-participle compounding are not sufficient to speak of grammatical constructionalization, it is merely constructional change. The discussion below will arrive at the same conclusion, but on the basis of a different argument.

While Traugott and Trousdale's definition of grammatical constructionalization as the emergence of a new node is very clear and intuitively plausible, it could be criticized for evoking the Sorites paradox, i.e. the question how many grains of sand it takes to make a heap. Just after how many constructional changes exactly do we have a construction that counts as a new node? The term, as defined, asks us to think of a discrete threshold. For the time being, such a threshold of grammatical constructionalization can only be identified after the fact, namely "*when constructs begin to be attested which could not have been fully sanctioned by pre-existing constructional types*" (Traugott and

Trousdale 2013: 22), that is, the occurrence of a construction in what is called a switch context (Heine 2002) or an isolating context (Diewald 2006). Since examples of this kind may be absent from the historical record or may occur only a long time after the point of constructionalization, it would actually be desirable to have a diagnostic of grammaticalization that does not hinge on the concept of switch contexts. The following paragraphs will thus develop a different criterion on the basis of which grammaticalizing constructions can be identified as such.

4.3 The upward strengthening hypothesis

A feature that grammaticalized constructions such as the *be going to* future construction, the *s*-genitive, or the ditransitive construction have in common is that their mental representations in speakers' minds are schematic. Some of their component parts, and actually all of them in the case of the ditransitive construction, are not lexically specified, but can be filled with a certain range of morpho-syntactic forms. Constructional analyses often use the concepts of 'slots' that accommodate structures of different kinds: a verb in the infinitive in the *be going to V* construction, a nominal group in the *s*-genitive, and so on. The emergence of constructional nodes with schematic slots is what Traugott and Trousdale (2013) capture with the term grammatical constructionalization. It is argued here that the emergence of schematic patterns is certainly an important part of the picture, but that there is another, equally important part, namely the process by which such patterns are cognitively strengthened once they are established. In other words, the grammaticalization of a linguistic unit is not only its initial constructionalization, but also its successive entrenchment. To illustrate this, Traugott and Trousdale (2013: 209) discuss how the pattern *a lot of X* constructionalizes into a quantifier construction. Switch contexts that provide evidence for its constructionalization are examples in which the open slot accommodates an abstract noun such as *truth* and examples in which the construction is used coreferentially with a plural pronoun (*a lot of sheep – they*). Through the repeated experience of such and other instances, the schematic representation of the construction is gradually strengthened, until the construction is fully grammaticalized. This, of course, is a point that Traugott and Trousdale do not deny and that has been argued independently for a long time (e.g. Bybee 2006: 717). What is claimed here, however, goes beyond that established consensus. The claim is that the gradual strengthening of schematic representations is something that only ever happens in grammaticalization. This idea is formulated more precisely in (6):

- (6) When the experience of a linguistic unit strengthens not only a mental representation of that unit itself, but also a mental representation of a more abstract construction, that process instantiates grammaticalization.

To give this idea a label, it will be referred to in the following as the upward strengthening hypothesis. Grammaticalization happens when the activation of a node in a constructional network causes a strengthening of a node that is situated at a higher, more abstract level of that network. Importantly, this idea is not an attempt to reduce grammaticalization to upward strengthening. The latter is a diagnostic of grammaticalization, not an exhaustive characterization. What is hypothesized is that grammaticalizing constructions receive upward strengthening, and that all others do not. The implications of that hypothesis are fleshed out below.

A core idea of usage-based approaches to language is that the experience of a linguistic unit will always strengthen the mental representation of that very unit. This process results in frequency effects such as ease of processing, reduction, and resistance to regularization (Bybee 2006). To illustrate, frequent phrases such as *I don't know* or lexical items such as *family* are swiftly processed and show phonological reduction effects; an idiomatic phrase such as *I kid you not* displays a resistance to word order changes that have taken place elsewhere. This kind of cognitive strengthening, often called entrenchment, can be contrasted with upward strengthening, since the strengthening affects just the linguistic unit that is heard, not a more abstract representation of that unit. The upward strengthening hypothesis predicts that the repeated use of a phrase such as *I don't know* will not increase the entrenchment of the more abstract constructions that are instantiated by that phrase, notably the intransitive construction and the negation construction. The upward strengthening hypothesis further predicts that instances of fully grammaticalized constructions, or of constructions in which the grammaticalization process has stopped, cease to project upward strengthening. What is predicted is that, for example, a newly encountered instance of ditransitive *give* will not further strengthen the mental representation of the more abstract ditransitive construction. By the same token, and to come back to noun-participle compounding, a newly encountered instance of *computer-based*, or even an entirely new coinage such as *Stratocaster-based* (which refers to an electric guitar that is based on the model called *Stratocaster*) is predicted not to strengthen the noun-participle compound construction. Importantly, upward strengthening and strengthening processes towards other levels in the network are not mutually exclusive. Encountering a moderately rare instance of noun-participle compounding, such as for instance *rain-sprinkled* is likely to yield not only upward strengthening, but also strengthening of the specific compound it-

self, and perhaps also strengthening of its component parts, *rain* and *sprinkled*. It is further necessary to point out that strengthening processes of this kind need not be viewed as a matter of either-or. As will be explained below, it is more appropriate to think of strengthening as a gradual process.

The previous paragraph invites several questions. Why would it be that experiencing an instance of a construction does not strengthen its schematic representation? What evidence is there to suggest that forms such as *Stratocaster-based* do in fact fail to trigger upward strengthening? And conversely, what kinds of experience would lead to upward strengthening? Is the upward strengthening hypothesis empirically testable? These questions will be addressed in the following two sections.

4.4 Which forms trigger upward strengthening, which ones do not?

One piece of evidence for the claim that certain forms, even newly coined forms, do not trigger upward strengthening is the observation that noun-participle compounding has undergone increases in type and token frequency, whereas the ratio of hapax legomena and the ratio of low-frequency forms have not increased. New instances of noun-participle compounding are continuously being coined, but apparently those are not the right ones to strengthen the schematic construction. Why do they fail to do so?

An explanation that suggests itself has to do with the process of categorization. In order for upward strengthening to occur, a linguistic unit that is experienced has to be categorized as an instance of a more abstract construction. If this categorization is not performed, or only superficially so, no upward strengthening will take place. Failure to categorize may be due to one of three reasons. First, there is evidence that speakers do not categorize certain linguistic units in terms of more abstract constructions when such a categorization is prohibitively hard. Taylor (2012: 59) discusses the example of *much*, which defies a straight-forward classification into a traditional part-of-speech category. In a sentence such as *We came to much the same conclusion*, it occupies the position of an adverb (cf. *We came to approximately the same conclusion*). In other contexts, such as *Much has happened*, it behaves like a pronoun. Assuming two different lexical entries will not solve the problem, since *much* deviates in its distribution from either of the two categories. Given such distributions, upward strengthening literally has ‘nowhere to go’. The second reason has to do with the subpatterns of an abstract construction. In cases such as *Stratocaster-based*, the absence of abstract categorization is due to the presence of a highly entrenched, more concrete

subpattern of noun-participle compounding, i.e. the *NOUN-based* construction. Patterns like these can be viewed as mid-level constructional schemas. The presence of prominent mid-level schemas in a constructional network may thus prevent upward strengthening to the most abstract node. More specifically, the encounter of *Stratocaster-based* will project upward strengthening to the *NOUN-based* construction, which in turn however, might not project any further upward strengthening to the noun-participle compound construction. Another English construction with prominent mid-level schemas is the ditransitive construction, which for instance has a schema that encodes the prevention of a transfer, as in *My boss denied me a raise* or *They refused him the information he needed*. Processing examples of this kind is likely to strengthen this particular semantic variant of the ditransitive, but not a more general constructional schema. A third reason for the absence of categorization, and hence upward strengthening, would be high text frequency. Highly entrenched linguistic units such as *I don't know* or *How are you doing?* tend to be mentally represented as chunks, which means that speakers do not analyze them into their component parts, which would be an act of categorization. Hence, a highly frequent noun-participle compound such as *home-made* should project less upward strengthening than a form such as *oxen-yoked*.³ The main ideas presented in this paragraph can be summarized in the following three predictions.

- (7) When a linguistic unit has a distribution that does not allow an unambiguous categorization in terms of a more abstract construction, experiencing that unit will not trigger upward strengthening.
- (8) When a subpattern of a schematic construction (a mid-level schema) is strongly entrenched, experiencing an instance of that mid-level schema will not trigger upward strengthening to the more abstract schematic construction.
- (9) Experiencing a linguistic unit that occurs with high text frequency will trigger less upward strengthening to its schematic construction than experiencing a linguistic unit that occurs with lower text frequency.

3 Importantly, categorizing tokens of usage as instances of constructions is not a matter of conscious reflection. The unconscious categorization of linguistic elements is evident for instance in syntactic parsing. The phrase *an old man's pyjama* can be understood as either 'a pyjama owned by an old man' or 'an old pyjama for a man'. While the respective meanings are available to consciousness, hearers are not conscious of the syntactic constituents that they form during parsing, nor of the fact that the two different interpretations correspond to two different constituent structures.

What still needs to be addressed is what kinds of linguistic units would trigger upward strengthening. In the context of what has been discussed above, upward strengthening would be triggered most strongly by linguistic units that force the hearer to engage in the process of categorization with a lot of cognitive dedication (although not at a conscious level). In order for this to happen, the linguistic unit should display two characteristics. First, it should provide an unambiguous cue to its category, so that categorization is facilitated. Second, it should be a marginal member of its category, so that its experience together with the cue is highly unusual, and therefore informative to the hearer. The following examples illustrate linguistic units of this kind.

- (10) a. inflation's consequences
 b. Not one man in five hundred could have spelled his way through a psalm.
 c. It is doing the dishes that I dislike most.

All of these examples are clearly identifiable as more abstract constructions, namely the *s*-genitive construction, the *way*-construction, and the *it*-cleft construction. And while the examples may not strike a present-day reader as particularly unusual, they would have been seen as that during earlier stages of English. The *s*-genitive has only recently come to be routinely used with inanimate possessors (Szmrecsanyi 2010), the verb *spell* is a late addition to the *way*-construction (Israel 1996), and *ing*-clause focus phrases in *it*-clefts are unattested in Middle English (Patten 2010). The requirements on linguistic units that are assumed to project upward strengthening are summarized in (11).

- (11) When a linguistic unit can be unambiguously categorized as a more abstract construction and it represents a marginal member of that category, experiencing that unit will trigger upward strengthening.

This requirement is in accordance with the observation that grammatical constructions do at some point stop to grammaticalize. Encountering an example of the ditransitive construction with *give* does not trigger upward strengthening because a ditransitive with *give* is not a marginal member of the category. The same applies to many hapax legomena of noun-participle compounding from the COHA, for instance the example shown in (12).

- (12) I could see tin cans and bottles sticking out from under a couple of moth-chewed buffalo robes on the ground.

Even though few readers will have encountered the word *moth-chewed* before, it should fail to trigger upward strengthening because it is not a marginal member

of its category, and hence does not force the hearer to re-consider the boundaries of that category.

4.5 Towards a test of the upward strengthening hypothesis

One important issue has been bracketed so far, namely how the upward strengthening hypothesis could be put to the test. A full empirical test of the hypothesis is beyond the scope of this paper, but this section will discuss how such a test could be attempted in principle. The key strategy of testing the upward strengthening hypothesis would be to re-create actual historical developments, as observed in diachronic corpus data, with a model that represents a dynamically changing constructional network. Computational networks of this kind have been described for instance in van Trijp and Steels (2012). Such networks represent linguistic units as nodes that are interconnected and hierarchically ordered. The network structure is dynamic and can change, which is in practice accomplished through simulated events of language use. Every time a linguistic unit is used successfully, that unit is ‘rewarded’, that is, its relative strength of representation vis-à-vis potential competitor units is strengthened (van Trijp and Steels 2012: 6). Simulations that involve many repeated runs of usage events suggest that this kind of strengthening process explains how populations of speakers converge on a shared lexical vocabulary. What is less clear, according to van Trijp and Steels (2012), is how such convergence is achieved with grammatical constructions that consist of multiple linguistic units. In order to explore how speakers converge on a common inventory of constructions, van Trijp and Steels (2012: 13) compare simulations in which usage events have different consequences in terms of strengthening. Specifically, they compare models in which hearing a multi-word sequence (such as for example *I don’t know*) strengthens (i) only the node that represents that very sequence, (ii) the node that represents the sequence, but also the nodes that represent the components of the sequence (i.e. *I*, *don’t*, and *know*), (iii) the node that represents the sequence but also more complex nodes that contain *I don’t know* as a subpart, and (iv) the node itself, its component parts and any overarching constructions, which is essentially a combination of (ii) and (iii). Simulation runs show that the fourth strategy, i.e. strengthening towards multiple levels in the network, yields the fastest rate of convergence in a population of linguistic agents (van Trijp and Steels 2012: 14–16). Strengthening processes between constructions at different levels in the framework thus facilitate the emergence of a shared grammar. While van Trijp and Steels conclude that strengthening towards multiple levels can explain how systematicity in a shared inventory of grammatical constructions arises in a population, they are careful to

acknowledge that actual languages are not perfectly systematic: They contain many partial generalizations and subregularities (2012: 11). In other words, the strengthening mechanism that is proposed may predict a level of systematicity that is never reached in actual languages. The upward strengthening hypothesis may serve as a constraint on simulation models in which strengthening is less omnipotent and which hence may show a greater tolerance for the pockets of irregularity that characterize natural languages.⁴

In practical terms, a constructional network representing noun-participle compounding would be based on the types and frequencies that are observable in corpus data, but beyond that, it would include abstract nodes and connections that are not directly observable. More specifically, for each historical period, the proposed network includes all attested types of a construction, with each type representing a node in the network. Each node has a strength value corresponding to its normalized text frequency. Using data from COHA, the first life-cycle of the constructional network represents the 1810s and includes types such as *blood-stained*, *fire-eyed*, or *heart-felt*, along with their normalized frequencies. Naturally, the nodes of the network are interconnected. Connections between the attested types are constructed on the basis of shared components, so that forms with the same participle (*awe-struck* <> *thunder-struck*) or the same noun (*heart-broken* <> *heart-felt*) are mutually connected. These links can be called horizontal connections, since they connect nodes that occupy the same level of abstraction in the network. On the basis of horizontal connections, schematic nodes such as *NOUN-struck* or *heart-PARTICIPLE* are stipulated as nodes at a higher level of the network. Nodes such as *awe-struck*, *moon-struck*, or *thunder-struck* would be connected to the *NOUN-struck* node via an upward connection. Finally, at the very top of the network there is a *NOUN-PARTICIPLE* node to which all unconnected nodes and all abstract nodes connect via upward links. Constructional networks of this kind are often referred to in the Construction Grammar literature (Diessel to appear), inheritance links and subpart links are firmly established concepts in that literature (Hilpert 2014: 60). However, some further elaboration of the network is necessary. Since the upward strengthening hypothesis distinguishes between marginal and central members of categories, this additional parameter would have to be implemented in the network, for instance via a semantic vector space model (Turney and Pantel 2010) that calculates a measure of distance between a type such as *saucer-shaped* and the

⁴ I would like to thank an anonymous reviewer for pointing me towards the discussion in van Trijp and Steels (2012) and its relevance for the upward strengthening hypothesis.

average of the entire network. Each node in the network would thus not only carry a frequency value, but also a value of semantic centrality.

Given such a network, it can be computed for every more abstract node (*NOUN-struck*, *heart-PARTICIPLE*, *NOUN-PARTICIPLE*, etc.), how much upward strengthening it receives if every lower node projects upward strengthening according to the stipulations in (7), (8), (9), and (11). The amount of upward strengthening towards a given abstract node constitutes the strength value of that node. Crucially, this value changes dynamically over time, as it is re-computed in every new period. As the constructional network changes from one period to the next, types appear and disappear, so that for instance *NOUN-struck* finds itself with fewer and fewer nodes that strengthen it. As semantic centrality plays a key role, the strength of abstract nodes is not a linear function of their type frequency, but rather a function of their semantic spectrum. Higher-level nodes can only gain in strength if new types appear that are semantically marginal. Since also text frequency plays a role, abstract nodes lose strength if their lower nodes become more frequent while everything else stays the same.

The main question of interest that can be investigated on the basis of such a network is whether the *NOUN-PARTICIPLE* node, at the very top of the network, becomes stronger over time. Since the upward strengthening hypothesis predicts that only grammaticalizing constructions should gain in strength, and since independent arguments suggest that noun-participle compounding has not recently undergone further grammaticalization, the implementation described above would prove the upward strengthening hypothesis wrong if the *NOUN-PARTICIPLE* node were to gain in strength over the COHA periods. Naturally, even if no increase in strength is registered, the picture would remain incomplete until a network of the same kind shows that in a grammaticalizing construction, the most abstract node does indeed gain in strength. Only if upward strengthening robustly shows itself across a range of grammaticalization phenomena, all the while staying absent from a range of other constructional changes, would the upward strengthening hypothesis be substantiated.⁵

⁵ An anonymous reviewer points out that in order to avoid circularity, it needs to be spelled out how the strengthening of a node is diagnosed and how a construction with a strong top node is predicted to behave. The diagnosis of strengthening is straight-forward: A gradual, numerically expressed amount of strengthening is projected upward from a newly attested instance of the construction according to the stipulations in (7), (8), (9), and (11). Strength can thus simply be ‘read off’ each node in the network, which is also done in studies such as van Trijp and Steels (2012). Strengthening can be diagnosed when the readings for a given node become larger over time. Nodes that gain in strength are predicted to show the behavior that Traugott and Trousdale (2013) describe as grammatical constructionalization. Increased productivity is only one aspect

5 Conclusions

This paper has pursued two objectives. The first has been an empirical investigation of how the word formation process of English noun-participle compounding has developed over the past two centuries. A central aspect of that investigation included the question whether the observable developments would correspond to developments that have taken place in the English *be*-passive. The main results of the empirical study can be summarized as follows. Noun-participle compounding substantially increases in token and type frequency while the ratio of hapaxes and low-frequency forms stays approximately constant. The recent decades show the rise of several highly prolific word families, notably involving the participles *based*, *related*, and *oriented*. None of the above developments are paralleled by changes that happened to the passive. This warrants the conclusion that noun-participle compounding is a construction that has been developing independently and that is mentally represented in its own right, even though it shares structural traits with the passive.

The second objective of the paper was to develop a theoretical understanding of these developments, which share certain traits with cases of grammaticalization but just as plainly deviate from them. Drawing on work by Traugott and Trousdale (2013), it was argued that the development of noun-participle compounding instantiates constructional change, not grammatical constructionalization. Based on that conclusion, it was argued that what is ‘missing’ in the case of noun-participle compounding is a process that was labeled upward strengthening. A diagnostic of grammaticalization in terms of relations in a constructional network was proposed, namely that grammaticalization happens when the activation of a node in a constructional network causes not only a strengthening of that node itself, but crucially also a strengthening of a node that is situated at a higher, more abstract level of the network. It was argued that upward strengthening can be observed in grammaticalizing constructions, but not in the use of lexical words or in the use of fully grammaticalized constructions. This proposal was called the upward strengthening hypothesis, and it was explained how that hypothesis could be put to the test by means of implementing its stipulations in a dynamically changing network. There is currently work in preparation that follows the steps that were described above.

To end this paper, it is perhaps in order to describe in a few words the potential benefits of approaches that implement hypotheses about grammaticalization

of this, the other two are increased schematicity and decrease in semantic compositinality. These can be assessed independently of productivity, so that the danger of circularity is avoided.

in dynamic models of language change. A first and fairly obvious point to make is that implementations of this kind can be a fast track to falsification. An idea, however plausible at first sight, may not work as predicted. Operationalizing grammaticalization in ways that allow measurements and quantitative assessments of hypotheses would be a contribution towards making grammaticalization a testable theory, which is a point that for instance Campbell (2001: 158) calls into question. Secondly, implementing processes of grammaticalization in constructional networks, however designed, and deriving insights from such implementations would show that Construction Grammar is more than a fashionable term in grammaticalization studies. The question whether Construction Grammar provides an added value to studies of language change is addressed occasionally (e.g. Hilpert 2013: 204), but since the idea of constructions has been implicit in many versions of grammaticalization theory, it would be necessary to explore its theoretical consequences in much greater detail. Modeling may be a way to do that, and to bring new constructional ideas into work on grammaticalization. Thirdly, thinking more in terms of computational modeling could establish links between the linguistic community that pursues research on grammaticalization and constructional change on one side, and the community of researchers that study language evolution with computational methods, specifically agent-based modeling (e.g., Wellens et al. 2013; van Trijp 2010, 2012) on the other. Making this link would not only yield a mutual enrichment of the two research communities, but it would also address a criticism made by Janda (2001: 266), who remarks that there is relatively little research that addresses “*those agents who are the real locus of language change: individual speakers and their collective social groups*”. To sum up these points, there is potentially quite a lot to gain, and even if the adoption of these new approaches will incur setbacks and the occasional interdisciplinary miscommunication, it seems like it is worth taking the trip.

Acknowledgments: I would like to thank two anonymous reviewers for very detailed and helpful comments. I hope that some of the more programmatic parts of this paper have become clearer through my engagement with those comments. I would further like to dedicate this paper to Günter Radden, who has recently celebrated his 80th birthday. Günter introduced me to Cognitive Linguistics when I was a student in his seminars on Metonymy and Cognitive Grammar at the University of Hamburg. Back then, I thought they were fun classes. It is only now that I can appreciate how much they influenced me. Thank you, Günter!

References

- Baayen, R. Harald. 2005. Morphological productivity. In Reinhard Köhler, Gabriel Altmann & Rajmund G. Piotrowski (eds.), *Quantitative linguistics: An international handbook*, 243–255. Berlin & New York: Mouton de Gruyter.
- Bauer, Laurie. 2006. Compounds and minor word-formation types. In Bas Aarts & April McMahon (eds.), *The handbook of English linguistics*, 483–506. Malden: Blackwell.
- Bauer, Laurie, Rochelle Lieber & Ingo Plag. 2012. *The Oxford reference guide to English morphology*. Oxford: Oxford University Press.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman grammar of spoken and written English*. Harlow: Pearson.
- Booij, Geert. 2010. *Construction morphology*. Oxford: Oxford University Press.
- Boye, Kasper & Peter Harder. 2012. A usage-based theory of grammatical status and grammaticalization. *Language* 88(1), 1–44.
- Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82(4), 711–733.
- Campbell, Lyle. 2001. What's wrong with grammaticalization? *Language Sciences* 23, 113–161.
- Davies, Mark. 2007. TIME Magazine Corpus: 100 million words, 1920s–2000s. Available online at <http://corpus.byu.edu/time/>.
- Davies, Mark. 2010. The Corpus of Historical American English (COHA): 400+ million words, 1810–2009. Available online at <http://corpus.byu.edu/coha>.
- Diesel, Holger. To appear. The network metaphor of usage-based construction grammar. In Ewa Dabrowska & Dagmar Divjak (eds.), *Handbook of cognitive linguistics*. Berlin & New York: Mouton de Gruyter.
- Diewald, Gabriele. 2006. Konstruktionen in der diachronen Sprachwissenschaft. In Kerstin Fischer & Anatol Stefanowitsch (eds.), *Konstruktionsgrammatik I. Von der Anwendung zur Theorie*, 79–103. Tübingen: Stauffenburg.
- Fabb, Nigel. 2001. Compounding. In Arnold Zwicky & Andrew Spencer (eds.), *The handbook of morphology*, 66–83. Oxford: Blackwell.
- Goldberg, Adele E. 1995. *Constructions. A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Extending collostructional analysis: A corpus-based perspectives on 'alternations'. *International Journal of Corpus Linguistics* 9(1), 97–129.
- Heine, Bernd. 2002. On the role of context in grammaticalization. In Ilse Wischer & Gabriele Diewald (eds.), *New reflections on grammaticalization*, 83–101. Amsterdam & Philadelphia: John Benjamins.
- Hilpert, Martin. 2013. *Constructional change in English: Developments in allomorphy, word formation, and syntax*. Cambridge: Cambridge University Press.
- Hilpert, Martin. 2014. *Construction Grammar and its application to English*. Edinburgh: Edinburgh University Press.
- Himmelman, Nikolaus. 2004. Lexicalization and grammaticization: Opposite or orthogonal? In Walter Bisang, Nikolaus P. Himmelman & Björn Wiemer (eds.), *What makes*

- grammaticalization – A look from its components and its fringes*, 21–42. Berlin & New York: Mouton de Gruyter.
- Hopper, Paul & Elizabeth C. Traugott. 2003. *Grammaticalization* (2nd edn.). Cambridge: Cambridge University Press.
- Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Israel, Michael. 1996. The way constructions grow. In Adele Goldberg (ed.), *Conceptual structure, discourse and language*, 217–230. Stanford: CSLI.
- Janda, Richard D. 2001. Beyond ‘pathways’ and ‘unidirectionality’: On the discontinuity of language transmission and the counterability of grammaticalization. *Language Sciences* 23, 265–340.
- Kay, Paul & Charles J. Fillmore. 1999. Grammatical constructions and linguistic generalizations: *The what’s X doing Y? construction*. *Language* 75(1), 1–33.
- Lehmann, Christian. 2002. *Thoughts on grammaticalization* (2nd edn.). Munich: Lincom.
- Lieber, Rochelle. 1983. Argument linking and compounds in English. *Linguistic Inquiry* 14(2), 251–285.
- Narrog, Heiko & Bernd Heine (eds.) 2011. *The Oxford handbook of grammaticalization*. Oxford: Oxford University Press.
- Noël, Dirk. 2007. Diachronic construction grammar and grammaticalization theory. *Functions of Language* 14(2), 177–202.
- Patten Amanda L. 2010. Grammaticalization and the it-cleft construction. In Elizabeth C. Traugott & Graeme Trousdale (eds.), *Gradience, gradualness and grammaticalization*, 221–243. Amsterdam & Philadelphia: John Benjamins.
- Plag, Ingo. 2003. *Word-Formation in English*. Cambridge: Cambridge University Press.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. New York: Longman.
- Sweetser, Eve. 1988. Grammaticalization and semantic bleaching. *Berkeley Linguistics Society* 14, 389–405.
- Szmrecsanyi, Benedikt. 2010. The English genitive alternation in a cognitive sociolinguistics perspective. In Dirk Geeraerts, Gitte Kristiansen & Yves Peirsman (eds.), *Advances in cognitive sociolinguistics*, 141–166. Berlin & New York: De Gruyter Mouton.
- Taylor, John R. 2012. *The mental corpus. How language is represented in the mind*. Oxford: Oxford University Press.
- Tomasello, Michael. 2003. *Constructing a language. A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Traugott, Elizabeth C. & Graeme Trousdale. 2013. *Constructionalization and constructional changes*. Oxford: Oxford University Press.
- Trips, Carola. 2009. *Lexical semantics and diachronic morphology. The development of -hood, -dom and -ship in the history of English*. Tübingen: Niemeyer.
- Turney, Peter D. & Patrick Pantel. 2010. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research* 37, 141–188.
- van Trijp, Remi. 2010. Grammaticalization and semantic maps: Evidence from artificial language evolution. *Linguistic Discovery* 8(1), 310–326.
- van Trijp, Remi. 2012. The evolution of case systems for marking event structure. In Luc Steels (ed.), *Experiments in cultural language evolution*, 169–205. Amsterdam & Philadelphia: John Benjamins.

- van Trijp, Remi & Luc Steels. 2012. Multilevel alignment maintains language systematicity. *Advances in Complex Systems* 15(3–4), 1–30.
- Wellens, Pieter, Remi van Trijp & Katrien Beuls. 2013. Fluid construction grammar for historical and evolutionary linguistics. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 127–132.
- Wiemer, Björn & Walter Bisang. 2004. What makes grammaticalization – An appraisal of its components and its fringes. In Walter Bisang, Nikolaus P. Himmelmann & Björn Wiemer (eds.), *What makes grammaticalization – A look from its fringes and its components*, 3–20. Berlin & New York: Mouton de Gruyter.